BERTering RAMS: What and How Much does BERT Already Know About Event Arguments? — A Study on the RAMS Dataset

Varun Gangal, Eduard Hovy
Language Technologies Institute
Carnegie Mellon University
{vgangal,hovy}@cs.cmu.edu

References

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

A Qualitative Examples

Some common error types we notice for BEST-HEAD are:

1. Errors where the argument is missed due to being in another sentence. This
2. Errors where the argument and gold argument are in the same coreference chain, but the argument picked is a different coreferent of the gold argument and not identical. We see this in Figure 4b in Experiments. We also see this in Figure 4.
3. Errors due to the same head being shared between co-occurring roles in the same example.
4. Errors due to pointing at adjectives/adverbs of the actual noun phrase. We see this in Figure 3 where the head points to *ambassador* in *ambassador Vitaly Churkin*, rather than *Vitaly Churkin*, which is marked as the gold argument.
5. Errors due to being distracted by metadata or extraneous name tokens, e.g reporter names. An example is Figure 9.
6. Errors due to being distracted by earlier occurrences of the same verb as the event trigger, like in Figure 11.

Through Figures 1 to 11, we present some additional qualitative examples to those in the main paper body. These include a mix of both successful argument identification as well as some failures to illustrate error types.

B Other Experimental Details

B.1 Stop Word List

For §2.5.3, the list of stop words we use is: I,you,he,she,we,they,them,our your,mine,my,their,theirs,ours and,or,along,with,beyond,under forward,backward,above,below,up,down who,what,which,how,when, where,much,it,its,upto,until,as, since,from,whose,whom,not

B.2 Training Details About LINEAR

LINEAR is trained for a maximum of 10 epochs, using an Adam optimizer (Kingma and Ba, 2014) with learning rate 0.01. Finally, the checkpoint chosen is the one with highest validation Acc (as defined in §2.3).

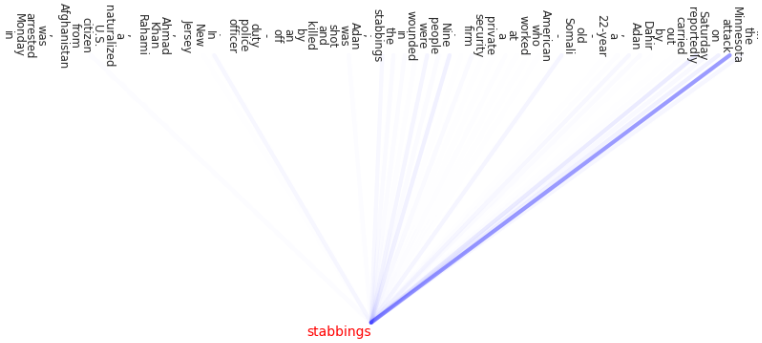

Figure 1: In this example, the head chosen by BESTHEAD for the PLACE role, correctly picks out the argument as "*Minnesota*".

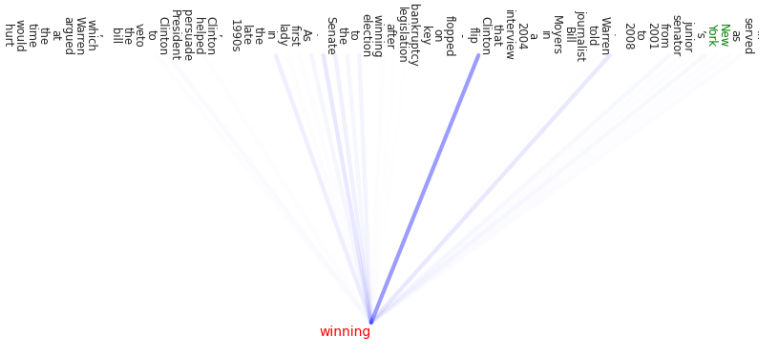

Figure 2: In this example, the head chosen by BESTHEAD for the PLACE role, incorrectly picks out the argument, confusing *Clinton* with *New York*

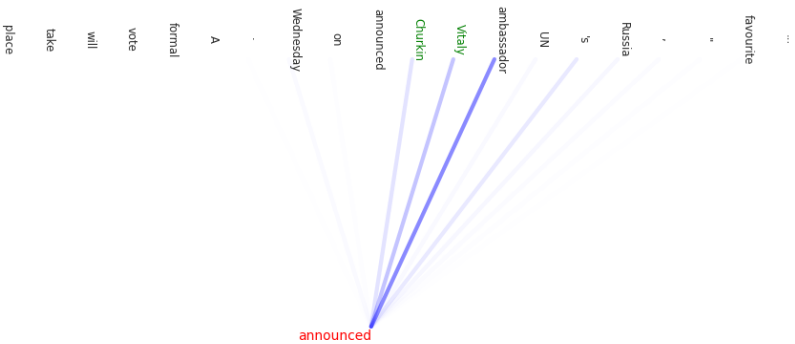

Figure 3: In this example, the head chosen by BESTHEAD for the COMMUNICATOR role, goes incorrect since it points to the adjective *ambassador* of the noun phrase *Vitaly Churkin*, rather than tokens in the noun phrase itself.

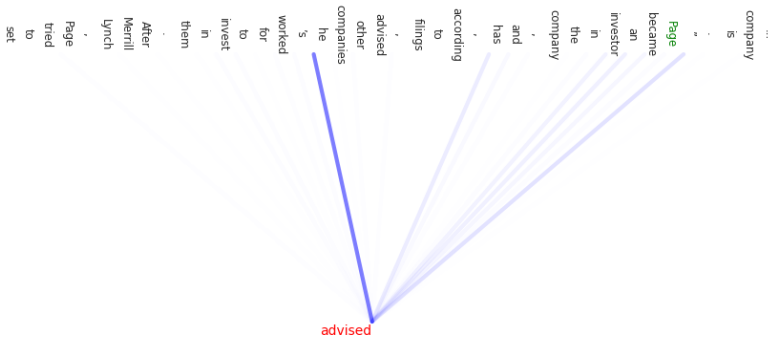

Figure 4: In this example, the head chosen by BESTHEAD for the COMMUNICATOR role, picks out the wrong coreferent (*he*) rather than *Page*, albeit from the correct coreference chain.

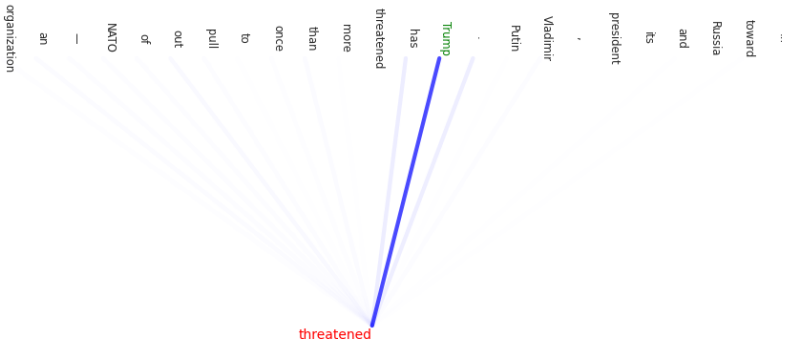

Figure 5: In this example, the head chosen by BESTHEAD for the PLACE role, incorrectly picks out the argument, confusing *Clinton* with *New York*

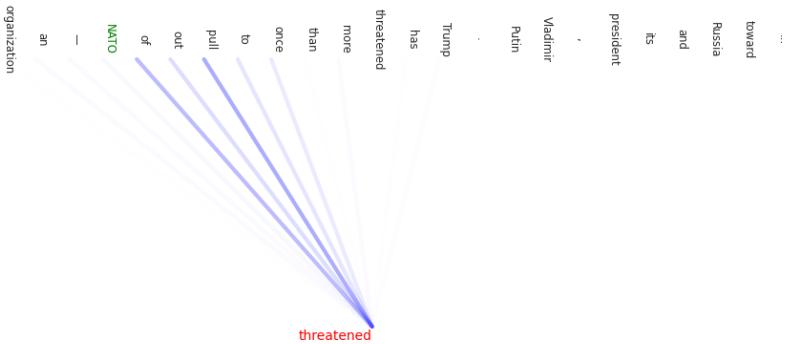

Figure 6: In this example, the head chosen by BESTHEAD for the RECIPIENT role, correctly picks out the gold argument *NATO*

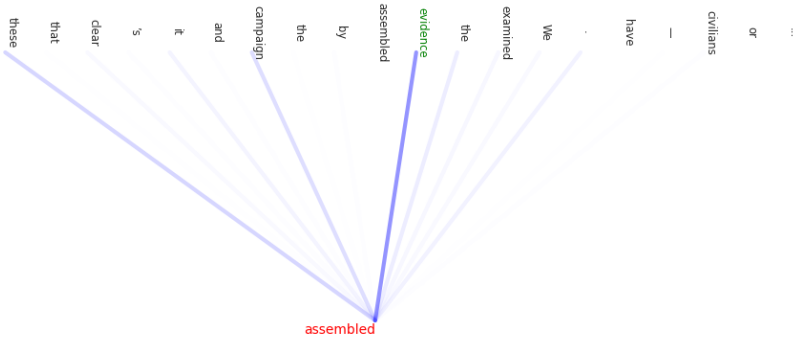

Figure 7: In this example, the head chosen by BESTHEAD for the ARTIFACT role, correctly picks out the gold argument *evidence*

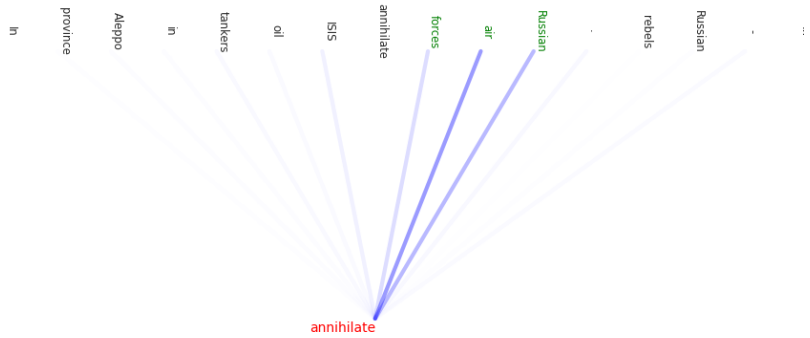

Figure 8: In this example, the head chosen by BESTHEAD for the ATTACKER role, correctly picks out the argument token *air* from the gold argument span *Russian air force*. Furthermore, we can see that the surrounding two tokens of the gold argument span, i.e *forces* and *Russian* are the second and third highest attention values in this head.

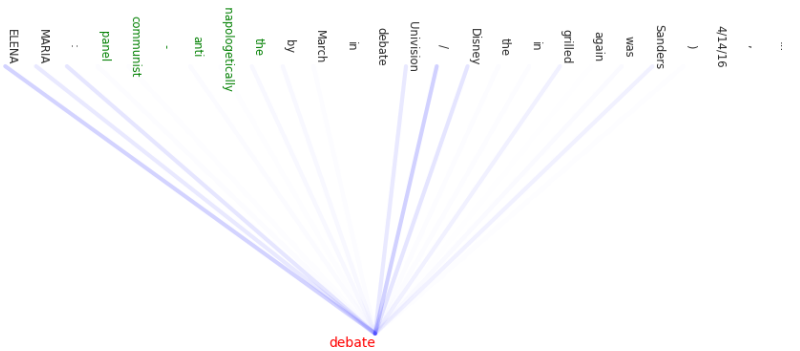

Figure 9: In this example, the head chosen by BESTHEAD for the PARTICIPANT role, incorrectly gets distracted by the reporter name *Maria Elena* extraneous to the article, missing the gold argument span *unapologetically anti communist panel*.

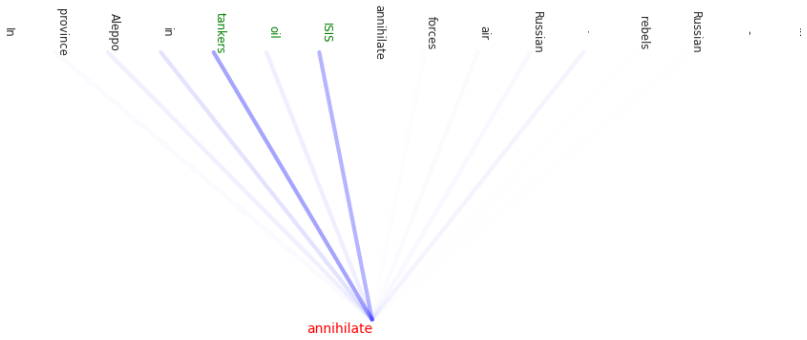

Figure 10: In this example, the head chosen by BESTHEAD for the TARGET role, correctly picks out an argument token *tankers* from the gold argument span *ISIS oil tankers*.

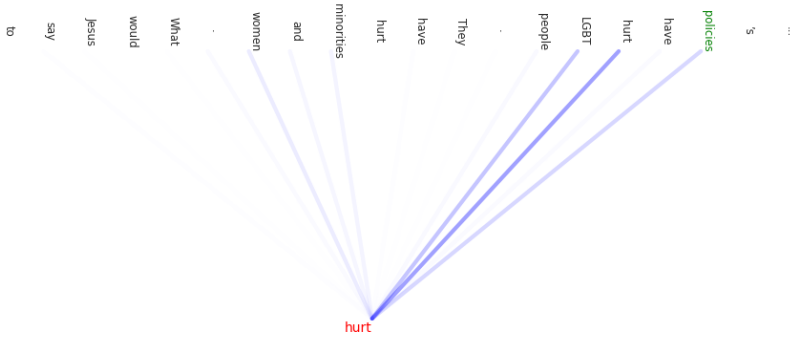

Figure 11: In this example, the head chosen by BESTHEAD for the INSTRUMENT role, incorrectly picks out an earlier instance of the trigger *hurt* instead of the gold argument token *policies*.

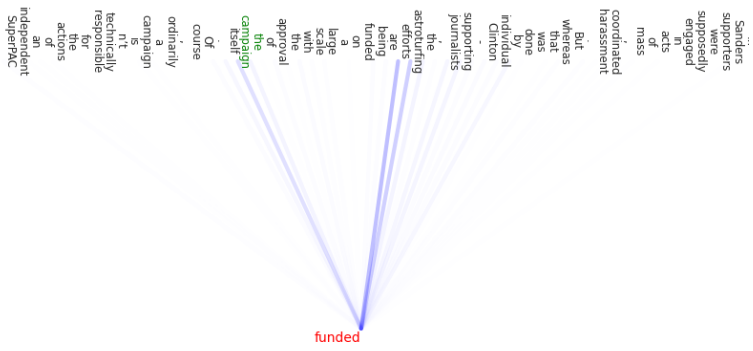

Figure 12: In this example, the head chosen by BESTHEAD for the BENEFICIARY role, incorrectly picks out a subevent *astroturfing efforts* referring to the full event *the campaign*, which is the most explicit *Beneficiary*. One may also call this a question of granularity.

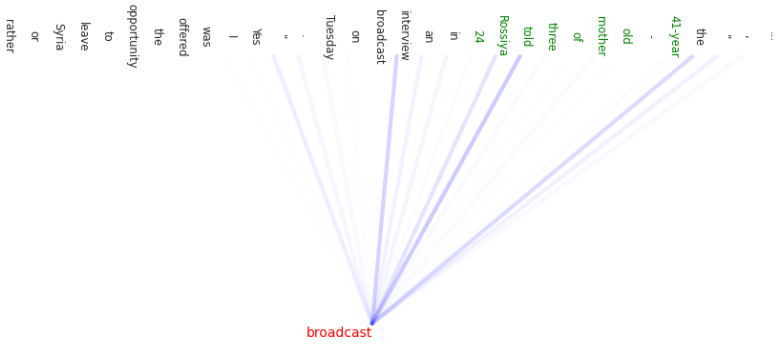

Figure 13: In this example, the head chosen by BESTHEAD for the COMMUNICATOR role, correctly picks out argument token *told* from the gold argument span *41-year - old mother of three told Rossiya 24*

B.3 Cross Sentence Performance: Additional Information

In Table 1, we record results for cross-sentence accuracy and the +CSO method for all roles

Role	BESTHEAD+CSO	LINEAR+CSO	CROSS-SENT %
DESTINATION	13.04 (21.43→0.00)	0 (39.28→0.00)	0
ORIGIN	4.76 (31.82→0.00)	4.76 (56.41→16.34)	31.82
TRANSPORTER	0.00 (31.58→0.00)	0 (43.42→0.00)	15.39
INSTRUMENT	47.37 (31.37→21.22)	52.63 (25.49→31.51)	37.26
BENEFICIARY	0.00 (26.56→0.00)	0 (34.37→0.00)	12.50
ATTACKER	12.50 (33.93→0.00)	0 (46.43→0.00)	14.29
TARGET	0.00 (44.61→21.71)	0 (44.61→0.00)	18.47
GIVER	7.14 (25.55→0)	0 (32.22→0.00)	15.56
VICTIM	49.99 (46.34→0)	0 (68.29→0.64)	7.32
ARTIFACT	22.22 (50.42→17.53)	11.11 (58.82→15.81)	7.57
COMMUNICATOR	19.99 (51.61→0)	9.99 (63.71→15.83)	8.07
PARTICIPANT	24.99 (28.57→6.37)	29.16 (30.72→6.37)	17.14
RECIPIENT	9.99 (40.78→0)	0.00 (44.69→0)	5.59
PLACE	15.31 (17.77→10.18)	30.61 (31.93→9.30)	29.52

Table 1: Accuracies on cross-sentence test examples using BESTHEAD+CSO and LINEAR+CSO. The values $Acc_{Total} \rightarrow Acc_{Cross}$ in parentheses are the total test accuracy and cross-sentence test accuracy respectively, using the simple version of the same approach i.e BESTHEAD and LINEAR.